

Strategic sustainability and financial performance: exploring abnormal returns

Janick Christian Mollet · Urs von Arx · Dragan Ilić

Published online: 30 April 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The ongoing empirical debate about whether SRI is associated, if anything, with subpar or surpassing financial performance is characterized by a somewhat indistinct focus and the infeasibility of tapping the full potential of existing models. By indistinct focus, we mean an analysis based on an aggregation of a myriad of SRI factors that potentially affect a firm's financial performance. The inability of taking full advantage of existing models is reflected by the fact that studies with European data have not been able to comprehensively account for systematic risk tilts. This paper presents a portfolio analysis that overcomes these issues by analyzing a distinct selection of small and innovative firms. We argue that both their strategic implementation of Corporate Social Responsibility and the general growth in socially responsible investments (SRI) lend themselves to an explanation for positive abnormal returns of this portfolio. We account for the idiosyncratic investment style of SRI by introducing a comprehensive pan-European risk-adjusted portfolio analysis based on the Carhart four-factor model. A novel

J. C. Mollet · U. von Arx · D. Ilić (✉)
University of Zurich, Center for Corporate Responsibility and Sustainability,
Zähringerstrasse 24, 8001 Zurich, Switzerland
e-mail: dragan.ilic@ccrs.uzh.ch

J. C. Mollet
e-mail: janick.mollet@ccrs.uzh.ch

U. von Arx
e-mail: urs.vonarx@ccrs.uzh.ch

J. C. Mollet · U. von Arx
ETH Zurich, Center of Economic Research, Zürichbergstrasse 18, ZUE F-13,
8032 Zurich, Switzerland

D. Ilić
University of Basel, Faculty of Business and Economics,
Peter Merian-Weg 6, 4002 Basel, Switzerland

propensity score matching method in conjunction with the estimation of structural models completes the conventional robustness checks in the literature.

Keywords Portfolio analysis · Asset pricing · Propensity score matching · Socially responsible investing (SRI)

JEL M14 · G11 · G12

1 Introduction

Despite the recent economic crises, socially responsible investing (SRI) has been strikingly resilient and not merely a caprice of prosperous economic times. SRI describes an investment strategy that accounts for environmental, social, and governance factors (ESG). Particularly in Europe, its expansion in the last decade has been impressive: Whereas in 2002, the European Sustainable Investment Forum reported 34 Bio Euro invested in so-called core SRI, for 2009 the same organization documented investments of 1,200 Bio Euro (Eurosif 2003, 2010). This reflects an annual growth rate of more than 66 %. The Sustainable Investment Forum for Germany, Austria, and Switzerland estimates that from 2005 to 2009, SRI assets in mutual funds, mandates, and other financial products have increased by roughly 29 % per year (FNG 2011). In comparison, for the mature US market, the Forum for Sustainable and Responsible Investment points out a 2.5 % increase in total assets managed under policies that incorporate ESG criteria from 2001 to 2010, from 2,010 Bio to 2,512 Bio US Dollars. This stands in contrast to US data from 1993 to 2001, when the same growth rate amounted to 43 % (US-Sif 2010). Even discounting the fact that a fair amount of the worldwide growth is attributable to bond investments and increasing asset coverage, it is fair to assume that in Europe, SRI in equities has grown substantially in the last decade, highlighting the economic importance for a systematic understanding of the financial ramifications that SRI entails.

But how might SRI affect the financial performance of a firm? Although conventional economic theory would roughly point to a negative association (Friedman 1970), recent and more pertinent descriptions of the phenomena give reason to believe that there is something to be gained by SRI. Due to its roots in principal-agent-theory, corporate governance is likely the most established theoretical branch in this regard. The significance of corporate governance in the context of shareholder and stakeholder value is discussed thoroughly by Tirole (2001). While the concept of stakeholder value—a concept which incorporates many ESG factors—fares worse than shareholder value on many levels (such as fuzzy tasking), it has conceptual advantages as well. Biased decision-making in presence of negative externalities, for instance, makes for one of the unintentional consequences of shareholder value. From a more technical perspective, SRI has been associated with a lack of optimal portfolio diversification for investors. Any restriction to the market portfolio is said to edge away from the efficient frontier. Yet Moskowitz (1972) counters that markets might fail to price value-relevant ESG factors, leading to higher abnormal returns for corresponding firms. In a seminal paper, Merton

(1987) presents an asymmetric information model in which investors overlook certain stocks. The smaller investor base is associated with suboptimal risk sharing, yielding positive abnormal returns for the stocks in question. In the next section, we take up this theoretical discussion; but ultimately, the question of whether SRI affects firm performance negatively or positively remains an empirical one.

There is an established literature examining the empirical effect of SRI on financial performance. Three methodologically distinct branches have evolved: Regression analyses, event studies, and portfolio analyses. Formerly a frequently used approach, simple regression analyses examine long-term effects of ESG factors on financial performance. They are, however, prone to omitted variable bias and reverse causality. Ambec and Lanoie (2007) summarize the results of nine such studies and conclude that there is limited evidence for a relationship.

In contrast, event studies mitigate the aforementioned caveats by focusing on a narrow time frame around an unexpected incident, say, an oil spill or employment layoffs. This firm-specific announcement is assumed to force a sudden reassessment of the firm's value on the part of its investors. In calculating abnormal returns within the specified time frame, significant market reactions to the announcement can be deduced. Many studies indicate substantial impacts on firm value from such incidents (Hamilton 1995; Dasgupta et al. 2001; Karpoff et al. 2005; Capelle-Blancard and Laguna 2010; Linn 2010). Although causally instructive, it remains unclear whether the observed market reactions have a lasting effect on firm value and why positive incidents often fail to show an impact.

Portfolio analyses, on the other hand, rely on a classic finance model in order to calculate a fund's long-term performance. Based on the capital asset pricing model (CAPM), the approach compares the abnormal returns of SRI funds with appropriate benchmarks. Of the three research branches, portfolio analyses currently constitute the most active one. Most studies gather data from actively managed SRI funds (see Derwall et al. (2011) for a comprehensive overview). So far, the empirical results are somewhat mixed, but, at least for the US, lean towards one direction. Hardly any of the studies indicate that SRI results in significant underperformance of the corresponding funds. Most studies cannot reject the assumption of equal performance. Indeed, some studies suggest that SRI funds fare better than the market.

Despite increasing sophistication, the current batch of portfolio analyses reveals some drawbacks of its own. By arbitrarily lumping together SRI funds that cater to different ESG priorities, specific ESG factors that are linked to financial performance are bound to drown in the aggregate of a myriad of non-relevant factors. In addition, the analysis of actively managed funds might distort the effect of the specific ESG factors by virtue of the portfolio managers' skills, transaction costs, or in- and outflows of assets. Most importantly, many studies disregard the risk tilts that are associated with the idiosyncratic investment style of SRI. It has been repeatedly shown that the risk factors introduced by Fama and French (1993) and Carhart (1997) explain substantial amounts of abnormal SRI fund returns (Bauer et al. 2005; Derwall et al. 2005; Bauer et al. 2007). Neglecting these factors bears the risk of false conclusions with respect to the effect of ESG factors on financial performance.

As yet, there have only been comprehensive data on US risk factors, which partly explains the heavy US focus of the literature. However, there are good reasons to

extend the field of research to Europe. Whereas in the US, the SRI market has evolved and expanded in the past decades and has settled somewhat by now, Europe is currently catching up (Bauer et al. 2005). Striking growth rates support this view. We argue that these demand shifts might be key in explaining abnormal financial performance. Europe and the US differ in other relevant aspects as well. Cortez et al. (2012) stress that there are geographical differences in the investment style of socially responsible funds. Data on market interest in nonfinancial information are consistent with such distinct priorities. Eccles et al. (2011a) show that US investors reveal a pronounced interest in governance data, while European investors are disproportionately keen on environmental data. This seems partly explainable by the fundamental differences in jurisdictions and political agendas. Lundgren and Olsson (2010) provide an example for different market reactions between the US and Europe using an event study: In Europe, environmental incidents are associated with a loss of value for the firm in question, while results are not statistically significant for US firms. Last but not least, extending the focus to European data meets the unanimous advice for out-of-sample evidence (Bauer et al. 2007; Derwall et al. 2005, 2011).

This paper investigates how a distinct selection of small European growth SRI firms fares in comparison to the market. In doing so, our portfolio analysis follows up the existing branch of studies and contributes to the literature by addressing the aforementioned problems along four dimensions.

First, we circumvent the distortion of an active portfolio management and the problem of an indistinct focus of aggregated funds by analyzing specialized firms selected by the sustainability research of the Zurich Cantonal Bank (ZKB). These firms, mainly small growth stocks with a focus on strategic Corporate Social Responsibility (CSR) innovation, constitute a ZKB internal universe from which the portfolio managers can pick. We treat this pool as a synthetic portfolio and compare it to the market.

In our second contribution, we introduce an exhaustive risk-adjusted portfolio analysis with European data. Comprehensive pan-European risk factors, constructed by Schmidt et al. (2011) have enabled this application study and account for the risk tilts that are rooted in the specific investment style of our SRI portfolio.

Third, in addition to the established errors-in-expectations argument for the explanation of positive abnormal returns, we discuss an overlooked structural argument based on rising demand in the SRI market.

Finally, our results prompted us to go one step further and check the robustness of our estimates in new ways. Using a propensity score matching (PSM) method, we match our portfolio annually to all the firms in the market and construct control portfolios with highly similar characteristics in terms of size, book-to-market value, and momentum; key figures upon which the risk factors are based. In addition, we account for country of origin and industry. To our knowledge, this method has never been applied in a portfolio study in combination with asset pricing models, introducing a novel technique to the existing literature.

We find a robust financial outperformance for the analyzed portfolio across different scenarios. In terms of the Carhart four-factor model, the monthly outperformance amounts to 1.3 %. Although there is reason to believe that some model misspecification is present due to the portfolio's tilt towards small and

medium sized growth firms, the outperformance is reduced at most to 1 % when adjusting for this issue.

The rest of this paper is structured as follows. The next section discusses our hypothesis that addresses the relationship between SRI and financial performance. In Sect. 3, we lay out the methodology of the asset pricing models and the data that are used for the hypothesis tests in Sect. 4.¹ Robustness of the analysis is dealt with in Sect. 5 using the matching method. Section 6 concludes.

2 SRI and financial performance

This section presents two distinct arguments that explain why our SRI portfolio might exhibit abnormal returns. First, the small growth firms distinguish themselves strategically based on their innovative approach in tackling sustainable solutions to societal problems. In contrast to conventional, more universal CSR classifications, we argue that our portfolio rather fits the notion of so-called strategic CSR. The relationship between generic SRIs and financial performance has been studied widely. Newer evidence indicates that this attention might have dispelled any errors in expectations by the market (Bebchuk et al. 2011; Derwall et al. 2011). We are thus more likely to find these mispricings in the specialized strategic CSR niche of our portfolio. The second argument relates to the influence of the rapidly rising market share of SRI on stock prices. Positive screening could lead to market disequilibrium accompanied by positive abnormal returns. Although both arguments can be characterized by market imperfections, they tackle the issue from different angles. The reasoning gives rise to our hypothesis and testable implications, which are the subject of the next sections.

Economists have traditionally been wary about the proposition that socially benevolent investments are financially beneficial for firms. A prevailing objection is the incompatibility with the objective of profit maximization, claiming that diverting resources from this goal for CSR reasons is bound to make a firm worse off (Friedman 1970). Even if CSR were to have any value-relevance, firms would have adopted it long ago. The premise is that if market participants could act more efficiently by adjusting their actions, they would do so by assumption. But evidently, not all choices are optimal in the first place, for we observe how companies keep finding new ways to raise their competitiveness by lowering costs and expanding market shares in innovative ways. Though often associated with a constraint, there is reason to believe that CSR can be used strategically to such ends as well. Recent economic frameworks suggest that there does exist a financial case for so-called strategic CSR.

The seminal publications of McWilliams and Siegel (2001) and Baron (2001), both of whom originally coined the term strategic CSR, stress the role of product differentiation in this context. In line with the predictions of strategic CSR, Siegel and Vitalino (2007) show that firms selling experience goods are more likely to engage in CSR for strategic reasons. Bagnoli and Watts (2003) define CSR as the private provision of a public good and present a model in which firms compete for socially responsible consumers. Besley and Ghatak (2007) pick up on this definition

¹ Some preliminary results of this paper were published in a leaflet of the Zurich Cantonal Bank (2011).

and discuss potential advantages of CSR as a provision of public goods. A prominent summary of the theoretical strategic CSR literature has been brought forward by Porter and Kramer (2006, 2011). They subsume the concept of strategic CSR and dub it shared value, “policies and operating practices that enhance the competitiveness of a company while simultaneously advancing the economic and social conditions in the communities in which it operates” (Porter and Kramer 2011, p 67). In a nutshell, shared value or strategic CSR is both beneficial for the firm and the community it operates in.

Strategic CSR involves a break from the commonly defensive and naive application of CSR. According to Porter and Kramer, by and large companies have made use of CSR as a preemptive or a mitigating shield that protects them from potential accusations concerning their business practice when, in reality, they should focus on the interdependence of business and society. To this end, it is proposed that companies incorporate aspects of CSR in a non-fragmented, targeted way as to establish a business strategy. So instead of addressing every possible justifiable CSR concern, a company should analyze its competitive landscape and its strengths and weaknesses that overlap with specific CSR aspects. On this basis, it should employ its comparative advantage to create a unique position that distinguishes it from the rest of the market. In this sense, strategic CSR is in line with the traditional economic argument of profit maximization as the firm’s objective.

To the casual observer, strategic CSR can be mistaken for generic CSR as they share the same toolbox. But instead of applying these tools loosely at different market frontiers, strategic CSR integrates them into a coherent business strategy. In order to differentiate between the generic and strategic CSR, one needs to be both knowledgeable about a firm’s strategy as a whole and the economic advantages that strategic CSR entails in that case. This is no easy task, in particular because the concept has merely begun to take root. Unlike generic CSR, strategic CSR cannot (yet) easily be extracted from ESG data but requires skilled expertise, entailing considerable costs. In turn, these costs put into perspective any financial market advantage resulting from such research.

The incorporation of potentially value-relevant information from strategic CSR is thus likely to take some time. This argument is supported by the fact that recently, some CSR studies have failed to find the superior risk-adjusted returns they used to. For example, Derwall et al. (2011) show that in the United States the positive abnormal returns of firms with a high score in employee relations as measured by Kinder, Lydenberg, and Domini (KLD) diminish over time.² Derwall et al. argue that any errors in investors’ expectations regarding value-relevant information of CSR are bound to be temporary as the market moves along its learning curve. Indeed, incorporation of CSR information, not only in form of KLD scores, has become ubiquitous among investors (Eccles et al. 2011a). Consequently, any value-relevant information that generic CSR potentially used to proxy for would likely have been arbitrated by now, accounting for the disappearance of positive abnormal returns. By the same token, formerly positive abnormal returns from good corporate

² In the empirical SRI literature, the KLD rating is an established proxy for the measurement of various dimensions of CSR for a selection of firms.

governance seem to have vanished. The widely cited governance-based investment strategy put forth by Gompers et al. (2003) and modified by Bebchuk et al. (2009) show how the US market apparently failed to correctly price so-called “entrenching” governance provisions in the 1990s. However, the superior returns from holding firms with few such provisions (and shorting the ones with many) have disappeared. In a follow-up paper, Bebchuk et al. (2011) show that during the 2000s any abnormal returns ceased to exist with respect to good governance. Like Derwall et al. (2011), Bebchuk et al. (2011) attribute this disappearance to a mispricing that has disappeared once markets have learned how to price these provisions correctly. Bebchuk et al. underpin this argument by showing that during the 1990s corporate governance was associated with fundamental firm differences in drivers of corporate performance: Until the 2000s (stock) market analysts were more positively surprised by the earnings announcements of well governed firms than by those of poorly governed ones, producing positive abnormal returns in the amount of 69 basis points from 1990 to 1999. The disappearance of abnormal returns from 2000 to 2008 coincided with increased attention to corporate governance by the media and market participants. In both Derwall et al. and Bebchuk et al. markets have eventually adjusted to a pricing equilibrium; not instantly but gradually over time.

To sum up the errors-in-expectations argument, given the present state of ESG data, strategic CSR seems more difficult to identify than generic CSR. On the other hand, economic theory suggests that strategic CSR potentially harbors value-relevant information, information that the market might have yet to incorporate.

In addition to investors’ errors in expectations, there is a complementary argument that can explain positive abnormal returns. Given plausible assumptions, it turns out that rising demand in SRI can affect abnormal returns. In empirical studies based on stock market data, an outperformance could stem from two sources. Better than expected fundamentals of companies (and the ensuing stronger stock demand of investors) on the one hand, and rising stock demand without changes in the companies’ cash flows on the other hand. Both effects lead to higher stock prices. Relating to the latter effect, particularly in Europe managed assets following SRI strategies have grown considerably in the last decade. It can be argued that this growth itself has affected stock prices positively.

Like any active investment strategy, SRI assumes that rational asset pricing models do not capture all relevant factors. After all, it is the aim of these strategies to outperform the market. Under rational asset pricing models, investors should simply buy stocks according to their current market weights. Some SRI investors, however, will deviate from current market weights of stocks for financial reasons. They will overweight certain stocks if they observe attractive SRI features that proxy for overlooked value-relevance and conversely will underweight stocks that they deem to be too expensive in this regard. As a consequence, SRI investment will have an impact on stock prices and investments of firms if demand and supply curves are not fully elastic (Wall 1995). This could happen if rational pricing models have not captured all relevant factors or if not all market participants invest accordingly. Following this argument, Petajisto (2009) puts forth a model in which CAPM investors, active managers (such as SRI followers), and noise traders can coexist in equilibrium.

The effect of rising SRI demand is more pronounced if the supply of stocks is less elastic. Stocks with a high market capitalization enjoy higher analyst coverage and larger institutional ownership. It is therefore plausible to assume that the supply of large capitalization stocks is more elastic than the supply of small capitalization stocks. From this perspective, our portfolio of small growth stocks portrays a pronounced case of a relatively inelastic supply. Following this argument, rising demand from SRI investors will affect stock prices of these companies all the more, leading to a financial outperformance in form of positive abnormal returns.

There is a body of empirical papers which are consistent with this argument. They analyze SRI performance in the US, where growth in SRI assets was particularly strong in the 1990s. In general, most of these studies find that risk-adjusted SRI returns are comparable to those of conventional benchmarks. There are some studies, though, which present a positive relation between certain CSR criteria and financial performance. Derwall et al. (2005) show that a portfolio scoring high on eco-efficiency criteria outperforms a portfolio of low scoring firms by 3.3 % per year from 1995 to 2003. Kempf and Osthoff (2007) find that a portfolio with high social responsibility ratings from KLD outperforms a portfolio with low scores by up to 8.7 % per year over the period from 1992 to 2004. Eccles et al. (2011b) use a matched sample of 180 companies and document that, from 1993 to 2010, a portfolio of high sustainability companies reveals an annual abnormal performance of 4.8 % compared to a portfolio of low sustainability companies. And most recently, Derwall et al. (2011) find that a portfolio of positively screened stocks, highly rated on employee-relations, earned positive abnormal returns of 5.6 % per year from 1992 to 2002. The exponential growth of SRI assets in Europe in the last decade coincides with our empirical time frame, which is indicative that rising demand might have affected financial performance.

Taken together, we identify two channels through which positive abnormal returns can emerge. Strategic CSR can contain value-relevant information, which due to errors in investors' expectations can lead to positive abnormal returns in the respective stocks. Complementary, rising demand in SRI can produce positive abnormal returns brought about by market frictions. These two arguments motivate our hypothesis that particularly in times of rising demand, small and strategic CSR stocks are positively associated with financial performance, a hypothesis we empirically test for in the next sections.

3 Data and methodology

Innovation acts as one particular channel of strategic CSR. Product and process innovation require foresight, assessments about future market dynamics, and likely considerable investments in the present. The focus is not on short-term profit maximization but on long-term growth, all features commonly associated with sustainability. With this notion, the sustainability research team of ZKB identifies small and middle capitalized firms offering innovative products and services. We henceforth refer to this portfolio as “innovators”. These innovators emphasize sustainable solutions from a societal point of view against the backdrop of climate change and the demographic trend of ageing. The firms can be allocated to the

Table 1 No. of innovators and market values

Years	No. of firms	Average market value (Mio USD)	
		Innovators	European firms
2002	16	332	997
2003	17	595	1,299
2004	13	396	1,608
2005	17	777	1,648
2006	30	1,178	2,074
2007	45	2,288	2,364
2008	75	645	1,185
2009	77	875	1,183

following six domains: renewable energy, energy efficiency, mobility, natural resources, health, and education. We conjecture that the innovators fit both the strategic CSR argument and the rising demand argument.

Table 1 reports the number of firms in the innovator portfolio and the average annual market value compared to our entire universe of European firms. Apart from the years 2006 and 2007, the average size of an innovator is roughly half the market value of the average European firm included in the Thomson Reuters Datastream and Thomson Reuters Worldscope database. The number of innovators increases from 16 firms in 2002 to 77 firms in 2009. The portfolio shows a distinct tilt towards the industrial sector in general and towards energy related industries in particular, with the services sector playing only a minor role. According to the International Classification Benchmark (ICB), Table 2 shows that about two-thirds of the 290 firms operate in the “Oil and Gas” and “Industrials” industry, 42 in utilities, 17 in the technology industry, and 16 in basic materials. This highlights that the innovator portfolio is tilted towards certain industries. A look at industry subsector levels in Table 3 reveals that roughly a third of the innovators deals with renewable energy equipment, followed by clusters geared towards industrial machinery, alternative electricity, alternative fuels, and building materials and fixtures. Table 4 presents the distribution of firm-year observations across the European countries. German firms constitute roughly 60 % of the portfolio. We address both sector and country bias in our robustness checks in Sect. 5.

It is worth pointing out again that this selection of firms constitutes a synthetic portfolio, and not an actively managed one. By virtue of this synthetic construction, we avoid issues of performance distortions such as portfolio manager skills or in- and outflows, which are generally unavoidable when analyzing performances of (sustainability) funds.³

³ Transaction costs such as commissions, bid/ask spreads, and market impact influence fund performance as well. For that reason, transaction costs are often considered in the practical implementation of theoretical investment strategies in synthetic portfolios. In our case, the inclusion of transaction costs would somewhat reduce absolute and abnormal returns, but only marginally as 82 % of the firms remain in the synthetic portfolio until 2009. Therefore, this strategy would have a low turnover and negligible transaction costs.

Table 2 No. of innovators across industries

Industries	No. of firms
Basic materials	16
Consumer goods	6
Consumer services	1
Financials	12
Healthcare	3
Industrials	80
Oil and Gas	113
Technology	17
Utilities	42
Total	290

In the heyday of the CAPM, value weighting would have been the obvious answer to the question which weighting scheme to apply to our portfolio. But ever since the empirical shortcomings of the CAPM became known, other weighting schemes such as Bayesian approaches to estimation error, methods of moment restrictions, portfolio constraints, optimal combinations of portfolios, or equal weighting merit consideration as well. After evaluating 14 models in seven empirical datasets, DeMiguel et al. (2009) come to the conclusion that the equal weighting 1/N portfolio strategy (which they dub “naive diversification”) “should serve at least as a first obvious benchmark” (p 1948). In order to reduce the influence of idiosyncratic risk in combination with firm heterogeneity in terms of size, we have therefore opted for the equal weighting method as our base case for both the innovator and the control portfolios.⁴

In order to assess how the innovators perform, we build a synthetic portfolio. For the appraisal of the performance of both the innovator and the control portfolios, which we construct in our robustness checks in Sect. 5, we employ pan-European risk factors of the standard four-factor CAPM based on Carhart (1997) as control variables. These risk factors have only recently been calculated by a group of researchers of the Universities of Aarhus and Zurich and ETH of Zurich (Schmidt et al. 2011).⁵ Thomson Reuters Datastream and Thomson Reuters Worldscope are the underlying databases for the calculation of these factors. The dataset consists of 43,005 European firms for the period of 1980 to April 2009. Both static and dynamic screens as well as numerous other quality checks reduced the number of firms available to 11,054 for the calculation of the market return, 9,462 for the calculation of the returns for size and value, and 10,035 for the calculation of the momentum factor, respectively. All financial data were converted into US dollars. For values before 2002, a fixed euro conversion rate was applied. Other currencies were converted via the at that time current exchange rate. The resulting database for

⁴ Calculations of value weighted portfolios tend to yield similar results and are available from the authors upon request.

⁵ By now, these factors also form part of the basis in the analysis of Mollet and Ziegler (2012), who investigate the performance of firms with high market values in the US and Europe.

Table 3 No. of innovators across subsectors

Subsectors	No. of firms
Alt. electricity	32
Alternative fuels	16
Auto parts	4
Banks	8
Building mat. and fix.	15
Con. electricity	7
Containers and package	4
Dur. household prod.	4
Electrical equipment	3
Electronic equipment	3
Financial admin.	7
Forestry	9
Healthcare providers	2
Hotels	1
Industrial machinery	31
Industrial suppliers	1
Investment services	4
Marine transportation	1
Medical equipment	1
Oil equip. and services	2
Renewable energy eq.	100
Semiconductors	4
Software	1
Specialty chemicals	7
Telecom. equipment	12
Waste, disposal svcs.	8
Water	3
Total	290

the factors size, value, momentum, and the market encompasses the period from January 1985 to April 2009.

Since its introduction, a wide range of empirical studies have shown that the one-factor model based on the market model (Sharpe 1964) and the CAPM (Markowitz 1952) describe the cross section of stock returns insufficiently. In particular, small and value stocks reveal systematically diverging stock returns which cannot be predicted by the market model. Based on the evidence of these “anomalies”, Fama and French (1993) advocate a three-factor model, which, in addition to the excess return of the market, adds two factors that proxy for size and value. Jegadeesh and Titman (1993) add to this insight. They show that stocks which have performed exceptionally well during the last 12 months tend to continue doing so in the next few months. Carhart (1997) incorporates this additional momentum factor into

Table 4 No. of innovators across countries

Countries	No. of firms
Denmark	9
Finland	4
France	9
Germany	175
Greece	2
Italy	2
Netherlands	2
Norway	11
Spain	14
Sweden	2
Switzerland	26
United Kingdom	34
Total	290

Fama and French's three-factor model. This four-factor model has proven to be better in terms of explanatory power of cross sectional stock returns and is now the most commonly used asset pricing model in financial economics.

The following equation formalizes the four-factor model:

$$r_{it} - r_{ft} = \alpha_i + \beta_{i1}(r_{mt} - r_{ft}) + \beta_{i2}\text{SMB}_t + \beta_{i3}\text{HML}_t + \beta_{i4}\text{WML}_t + \varepsilon_{it}$$

The excess return of a stock is decomposed on the basis of the four aforementioned factors market, size (SMB), value (HML), and momentum (WML). r_{it} stands for the return of a given stock portfolio in month t , r_{ft} is the return of the risk free asset (in our case, the euro 1-month libor rate), and r_{mt} is the return of the pan-European market. The return series for the factor SMB_t is the difference between the returns of portfolios consisting of stocks with low market capitalization and portfolios consisting of stocks with high market capitalization. The return series for HML_t is the difference between the returns of portfolios consisting of stocks with a high book-to-market ratio and portfolios consisting of stocks with a low book-to-market equity ratio. Finally, the return series for the momentum factor WML_t is the return difference between a portfolio of past winners and a portfolio of past losers. The application of the four-factor model yields estimates for the coefficients β_{i1} , β_{i2} , β_{i3} , β_{i4} , and α_i . The coefficients β_{i1} , β_{i2} , β_{i3} , and β_{i4} mirror the sensitivities of a given portfolio to the factors market risk, size, value, and momentum, ε_{it} stands for the error term, and α_i reflects the average periodical (positive or negative) abnormal performance which cannot be explained by the four factors.

4 Results

The average monthly market return over the entire portfolio time span from 1/2002 to 4/2009 amounts to 0.30 % (Table 5). This average return over 88 months is 0.01 percentage points below the yield of the risk free rate. When partitioning the full

Table 5 Average monthly returns over time in %

	r_{mt}	r_{ft}	Innovators _t	SMB _t	HML _t	WML _t
1/2002–4/2009	0.30(0.59)	0.31(0.01)	0.74(1.06)	0.41(0.24)	0.55(0.15)	1.12(0.45)
1/2002–8/2005	1.37(0.67)	0.23(0.00)	1.40(1.38)	0.80(0.31)	1.05(0.24)	1.08(0.68)
9/2005–4/2009	−0.78(0.95)	0.39(0.02)	0.08(1.63)	0.03(0.37)	0.05(0.14)	1.16(0.59)

Average geometric returns are displayed and standard errors are shown in parenthesis. Value weighted returns are presented for the market return r_{mt} and the factors SMB, HML, and WML. Equally weighted returns are displayed for the Innovator portfolio. r_{ft} is the risk-free rate

time period into two equally long sub-periods, we see that the dismal market performance is attributable to the second sub-period covering 9/2005 to 4/2009. This sub-period captures the financial crisis triggered in 2007 and yields an average monthly market return of -0.78% , while the risk free rate manages to return 0.39% . On the other hand, in the first sub-period between 1/2002 and 8/2005, the market generated monthly returns of 1.37% , whereas the risk-free interest rate offered only 0.23% .

Focusing on the average monthly return of the innovators in Table 5, we observe that the yield of 0.74% is substantially higher than the market return and that this result is driven by the second sub-period, during which the innovators fared much better than the market. The risk factors offer a more nuanced insight of this performance. WML indicates that winners from the previous months by far outperform recent losers in all time periods: The momentum strategy generates monthly returns exceeding 1% in all periods. The returns of the SMB and HML strategy, 0.41 and 0.55% , respectively, are both explained by their strong performance in the first sub-period. In the second sub-period, these strategies were barely profitable. These observations are in line with the theoretical reasoning of SMB and HML. The second sub-period largely covers an economic recession. During these harsh times, small and value firms are said to perform worse. Small firms exhibit problems in raising capital, and value firms find themselves unable to substantially lower any running capital costs, something growth firms are apt to do. Unlike value firms, growth firms can postpone investment expenditures as they are more flexible.

Table 6 reports the Carhart four-factor model regression results for the full time period and the two sub-periods. Within a given estimation window, the factor loadings are assumed to be fix even though they might change as portfolio composition, risk perceptions, and demanded risk premia are likely to vary through time. In this sense, the sub-periods serve as a rough first robustness check in terms of temporal dependence. Below, we implement a more sophisticated approach to this issue. In order to control for possible distortions in the covariance of the estimates due to heteroskedasticity or autocorrelation in the disturbance term, only robust heteroskedasticity- and autocorrelation-consistent z-statistics according to Newey and West (1987) are reported. In line with common practice (Greene 2002), the error structure is assumed to be possibly autocorrelated up to three lags.⁶

⁶ Current practice specifies the number of lags to be approximately the fourth root of the number of time series observations.

Table 6 Parameter estimates and Z-statistics in the four-factor model for different time periods

	Alpha	$r_{\text{int}} - r_{\text{it}}$	SMB _t	HML _t	WML _t	R ²
01/2002–04/ 2009	1.30**(2.14)	1.51*** (10.94)	0.35(1.18)	-1.16**(-2.37)	-0.07(-0.30)	0.61
01/2002–08/ 2005	1.25(1.14)	1.32*** (4.60)	0.71** (2.04)	-1.48**(-2.12)	-0.26(-0.75)	0.45
09/2005–04/ 2009	1.54(1.66)	1.54*** (9.01)	0.22(0.66)	0.59(0.60)	0.15(0.56)	0.74

* (**, ***) means that the parameter is different from zero at the 10 % (5, 1 %) significance level, respectively. The parenthesis below the point estimates contain the z-statistic, corrected for heteroskedasticity and autocorrelation up to three lags

Table 6 shows the parameter estimates of the Carhart four-factor model. The alpha of the equally weighted innovator portfolio is outperforming the value weighted market monthly by 1.30 %, a return which is significant at the 5 % level. In contrast to the descriptive analysis, this outperformance cannot be clearly attributed to a particular sub-period. Because the market beta is always larger than one, the innovator portfolio is magnifying market movements and is therefore exposed to a higher systematic risk.

The significant and positive SMB loading in the first period highlights that the innovator portfolio performs well due to its tilt towards small stocks. On the other hand, the HML factor captures a significant negative loading variation of the innovator portfolio return in the full time period. This is attributable to the first sub-period because the point estimate of the HML loading in the second sub-period is low and not significant. Recall that Table 5 indicates that value firms outperform growth firms during the first sub-period, driving the outperformance over the full time span. The innovator portfolio mainly consists of small growth companies and exhibits a significant exposure to the risk factors SMB and HML. The outperformance of small cap stocks in the first sub-period therefore added to the absolute return of the portfolio. The strong tilt towards growth stocks on the other hand reduced the absolute performance in first sub-period when value stocks had a good run. But the outperformance of the innovator portfolio in comparison to the European market—as displayed by the statistically significant alpha of 1.30 % per month—was achieved more or less continuously over the whole observation period. As we will see in the following section, the growth character of the innovator portfolio is also supported when comparing the book-to-market-equity ratios in Table 9.

The two sub-periods offer only a limited glimpse into the intertemporal changes in portfolio composition, risk perceptions, and demanded risk premia. Figure 1 gives a more differentiated view by plotting the coefficients of the four-factor model and their 95 % confidence intervals for a moving window regression. The method uses the estimation technique from Table 6 and applies it to a rolling regression time window of 36 months. This window selection translates into 30 degrees of freedom in the model estimation. The first regression

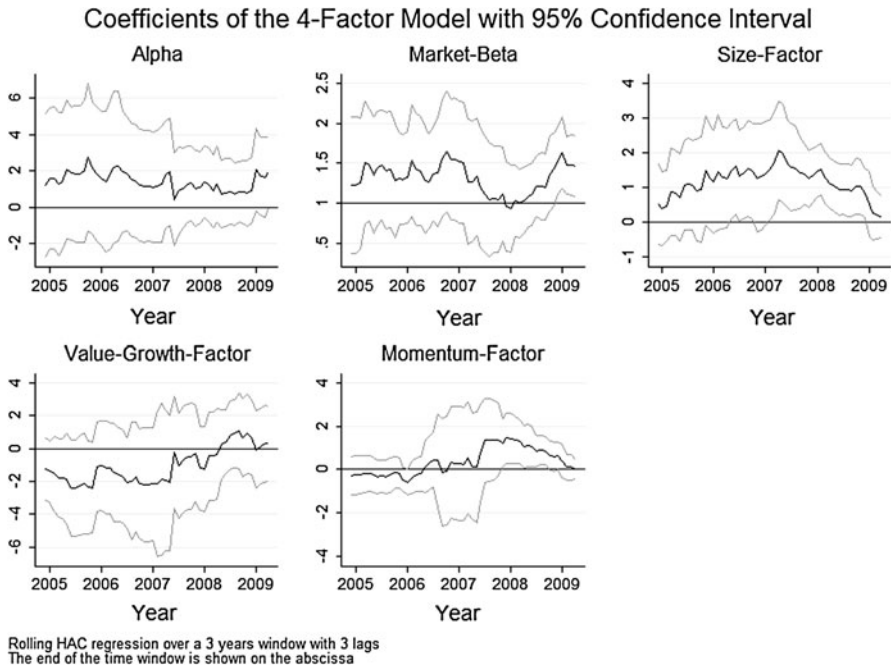


Fig. 1 Coefficients with 95 % confidence interval of rolling regression in the four-factor model

window covers 1/2002–12/2004. Its coefficients and confidence intervals are plotted at the end of this range. Subsequently, the plots are updated by moving the time window forward monthly.⁷ This technique has the advantage of not relying on disjoint and arbitrary time periods and allows for flexibility and structural interruptions in the estimates. The rolling regression plot for the alpha illustrates that the point estimates of the intercept of the innovator portfolio remain positive throughout the observed time window. The previous analysis of the two sub-periods revealed that the positive alpha is not driven by either time window. The rolling regressions are in line with this view, lending credibility to the robustness of the estimates.⁸

The results support our hypothesis. The portfolio of small growth firms with a strategic CSR implementation realizes a systematic stock market outperformance during 2002–2009. In terms of the hypothesis, this outperformance might be attributable to overlooked but value-relevant strategic CSR characteristics of these firms and to positive stock price effects resulting from a rapidly growing SRI demand. Due to the size of its stocks, the portfolio is especially prone to such stock price effects because stock price elasticity is likely positively related to firm size.

⁷ Due to different time windows, the graphical representation of the four-factor model in Fig. 1 is not directly comparable with the results in Table 5.

⁸ As an additional robustness check, we estimated the four-factor model with the both the market return and the risk factors computed on the basis of equally weighted returns. The results did not change.

However, there remains a methodological caveat associated with the findings. The results might not be reliable if an inappropriate model was applied to describe equilibrium asset prices. To clear up doubt on the reliability of the results, in the next section we evaluate the performance of the innovator portfolio compared to similar benchmark portfolios.

5 Robustness checks

5.1 A matching approach

Market efficiency can only be tested jointly with some equilibrium model, in this case an asset pricing model. Therefore, one might argue that the interpretation of anomalous evidence is ambiguous, because the hypothesis of zero abnormal returns is tested jointly with the hypothesis that the asset pricing model used to estimate abnormal returns is valid (e.g. Fama 1970, 1991). This so-called joint hypothesis problem is well known in the literature. It implies that if the used asset pricing model provides an imperfect description of expected returns, the estimated intercept (alpha) represents the combined effects of model misspecification and mispricing. Indeed, the Fama and French (1993) three-factor asset pricing model exhibits difficulties in explaining the average stock return of small firms and firms with high book-to-market ratios (e.g. Barber and Lyon 1997). In particular, Fama (1998) is concerned with bad model problems when considering equal weighted portfolios because all common asset pricing models have systematic problems in explaining the average returns of small stocks. The joint hypothesis problem thus seems to be of eminent interest when equal weighted stock returns of small to middle sized growth firms are analyzed. This motivates further robustness checks in our case.

Lyon et al. (1999) demonstrate that the three-factor model is misspecified for samples of non-randomly selected firms. They recommend comparing sample firms to the general population based on size (SIZE) and book-to-market equity ratios (BEME) as well as on the pre-event return performance and other characteristics. Likewise, Barber and Lyon (1997) favor this control firm approach when comparing different buy-and-hold abnormal return methods to evaluate long-run abnormal returns in event studies.

In order to investigate if the results in Sect. 4 are driven by bad model problems, we investigate whether we find similar anomalous evidence in portfolios with similar characteristics as the innovator portfolio. For one, the innovator portfolio is biased in SIZE and BEME. It also reveals substantial country- and industry-specific expositions. One might argue that some of these extreme portfolio characteristics produce extrapolation bias or that the linearly specified Carhart four-factor model might not be well specified to explain the returns of this portfolio. Ho et al. (2007) illustrate that model dependence in general can be reduced by combining the estimation of structural equation models with data preprocessing by matching treatment and control groups. The central idea of this approach is that if exact matching is not feasible, common parametric procedures can potentially improve

inference even after matching because of variation in the covariates. In this vein, we compare the asset pricing model estimates of the innovator portfolio to synthetic reference portfolios with similar characteristics. If these control portfolios exhibit similar abnormal returns in comparison to the innovator portfolio, we have evidence that the Carhart four-factor model might indeed suffer from misspecification for this portfolio type.

Portfolio matching methods have been applied previously in the literature. Mitchell and Stafford (2000), for example, consider SIZE and BEME in the construction of reference portfolios on the basis of specific portfolio assignments. In contrast to their approach we construct reference portfolios not by traditional matching methods but by matching the innovators to all European firms in the market based on a propensity score. Conventional matching approaches notoriously have difficulties in simultaneously matching multiple dimensions. Their sequential approach—such as prescreening based on SIZE, followed by a nearest BEME matching—implies that optimizing matching quality in one dimension usually comes at the cost of deteriorating matching quality in other dimensions. In contrast, PSM reduces a multi-dimensional problem to one single score, the propensity score (Rosenbaum and Rubin 1983).

Li and Zhao (2006) use PSM to identify similar firms based on size, BEME, and momentum. In studying the event-time buy-and-hold abnormal return approach of firms with primary seasoned equity offerings, Li and Zhao compare the mean and median returns of the event and the matched control firm portfolios, albeit without estimating any parametrical model. They find that the PSM method delivers better results than the conventional matching methods applied by Lyon et al. (1999), explaining away partially the anomalous behavior of small issuers. They also find that adding additional variables in the propensity score specification estimation has only marginal effects. In other settings, PSM has been applied by Aggarwal et al. (2009) to compare the governance of US and non-US firms and by Drucker and Puri (2005) to investigate the benefits of concurrent lending and underwriting. To our knowledge, PSM has yet to be applied in conjunction with asset pricing models to evaluate abnormal returns. We take up on the argument of Ho et al. (2007) and combine PSM with a structural model in order to check for model dependence.

In the estimation of the propensity score variables affecting the dependent and the treatment variable should be included, termed “minimum relevant information set” by Heckman and Navarro-Lozano (2004). On this note, following Li and Zhao (2006) we use firm level SIZE, BEME, and 11-month pre-matching momentum (MOM) for a 1-to-1 nearest neighbor PSM. We argue that using a PSM, we can construct an appropriate evaluation benchmark to reappraise the robustness of our results.

If a firm is identified by the ZKB research team as an innovator in year γ , the firm is retained in the innovator portfolio in the year $\gamma + 1$. This implies that the innovator portfolio is sorted in December of year γ and its holding period is $\gamma + 1$.

$$\text{Prob}(\text{Innovator} = \text{yes} | x_\tau) = \Lambda(x_\tau, \beta) \quad (1)$$

$$x'_\tau = \text{SIZE}_\tau, \log(\text{SIZE}_\tau), \log(\text{SIZE}_\tau)^2, \log(\text{SIZE}_\tau)^3, \text{BEME}_\tau, \text{MOM}_\tau \quad (2)$$

Each December, we estimate Eq. (1) with a logistic regression. If linear covariates in the propensity score estimation are not producing good matching quality, higher order terms or interaction terms should be added according to Rosenbaum and Rubin (1983). Starting with linear covariates, we added transformations of the SIZE variable to obtain better balance on this dimension, tried different other specifications, and concluded that the predictor variables in vector x_τ of Eq. (2) deliver the best balance as measured by mean and median differences for pooled PSM over all years.⁹ The index τ is a monthly indicator and fixed to December of each year in Eq. (2). SIZE_τ is the market value in million USD, the book-to-market ratio BEME_τ is the ratio of common equity to the market value of equity, and MOM_τ is the geometric average of the monthly returns over a window of 11 months from $t-12$ to $t-2$.¹⁰ Based on the vector x_τ , the probability of being an innovator firm is predicted for the general population of European firms. For the matching procedure, we adhere to the guidelines provided by Guo and Fraser (2009) and use the predicted probability p to define $\text{logit} = \log \frac{1-p}{p}$ as the propensity score. In turn, this score is used for a 1-to-1 matching without replacement in descending order with a caliper of one-fourth of the standard deviation of the annual propensity scores to ensure a certain level of common support. The matching procedure is done using the Stata package `psmatch2` from Leuven and Sianesi (2003). Nearest neighborhood matching is executed separately every year. In the base case scenario the nearest matches are restricted to the same countries because our result might have been confounded by the overrepresentation of German firms in our sample.¹¹ Due to this country restriction and the used caliper, innovator firms are being dropped in a specific year if no other firm within the same country has a sufficiently close propensity score. Control firms constitute firms in the market with monthly return values in the sorting year and with plausible available values of SIZE and BEME in December of the sorting year.¹² Because the control firm portfolio is held in the months after December of the sorting year, potential control firms need to remain in the market for the entire year after the sorting year. In addition to the country base case we construct two alternative control firm portfolios by restricting the nearest neighborhood PSM to the same industry and (in the most restrictive scenario) to both the same country and industry. The last matching scenario requires exact matching in two dimensions and inflates the number of cells considered.

⁹ The aim of the matching specification is to improve the balance between the innovator firms and the control firms with respect to the variables assumed to be of relevance. We tried different combinations and transformations of the firm characteristics SIZE, BEME, and MOM and identified the specification (2) as delivering the most balanced samples in our favored matching scenario.

¹⁰ Per common practice, common equity is required to be greater. In addition, we added investment tax credits to common equity.

¹¹ The Thomson Reuters Datastream Variable 'GEOGN' was used as the country identifier.

¹² Book equity is the sum of common equity and investment tax credit and is restricted not to be negative. In addition, we rely on the filters applied in Schmidt et al. (2011).

It consequently reduces the number of firms in each cell. Because of the caliper, this results in reduced sample sizes. Whereas our original sample contains 290 firm-observations, the exact country restriction and the exact industry restriction yield 237 and 236 firm-observations, respectively. The most restrictive exact country and industry match leaves us with 199 firm-observations.

Additionally, similar to Mitchell and Stafford (2000) we apply resampling procedures to derive the expected abnormal performance of the innovators given their sample composition. We still identify similar firms by PSM but now match, in each year, the ten most similar firms to every innovator firm. From this pool of control firms we draw 1,000 calendar-time portfolio samples and estimate the expected intercept from these random samples by sampling in each draw (with replacement) 30 control firms in every year. Based on these draws we also estimate the expected intercept of portfolios of innovators that are most similar to the sampled control firms. Applying different matching scenarios, we estimate alpha distributions conditional on differently composed sample compositions. First, we use PSM based on the characteristics SIZE, BEME, and MOM. We then restrict the ten nearest neighbor PSM to the same country and, in addition, to the same industry. Finally, we require an exact country and industry match. In this last matching specification, we have to weaken the matching specifications in order to obtain a sufficient sample size by matching only the five nearest neighbors and by dropping the caliper. The resampling approach is of interest because there might not be one specific firm that serves as the definite control firm. Moreover, the considered time period of 8 years is not very long but quite turbulent, making it more difficult to find significant results. Therefore, the inspection and comparison of the alpha distribution of resampled control and innovator portfolios might provide additional insight.

5.2 Results with control firms

Table 7 reports the number of firm-years in the innovator portfolio and the generated portfolio of control firms across different industries for the country match, our base case scenario. Among the industries, “Oil and Gas” and “Industrials” continue covering the major part of firm-years in the innovator portfolio. In contrast, the control firm portfolio reveals a tilt towards “Financials”. Table 8 verifies the effectiveness of the exact country match. In Tables 9, 10, and 11, the portfolio characteristics SIZE, BEME, and MOM of the innovator (I), control (C), and market (M) portfolios are juxtaposed. The means and medians of these portfolio characteristics offer a rough picture of the balance improvement achieved by the matching. Table 9 lists annual means and Table 10 reports annual medians for the three portfolios and the identified portfolio characteristics. On an annual basis, these tables confirm that the characteristics SIZE, BEME, and MOM are more similar between the innovator and the control firm portfolio than between the innovator and the market portfolio. These properties are summarized in Table 11. From 2001 to 2008, the mean absolute difference between the control and innovator portfolio on the one hand and the market and innovator portfolio on the other hand shows a tenfold reduction in the control portfolio for the characteristic SIZE and a threefold

reduction for BEME and MOM in comparison to the market. These absolute differences are averaged across all the years for the annual means and medians of the three characteristics. The absolute mean differences of the medians tell a similar story. The differences in SIZE, BEME, and MOM are two to seven times smaller for the control firm portfolio than for the market portfolio. The control firm portfolio thus appears to be a more appropriate benchmark for the innovator portfolio.

The average monthly return of the market and the risk-free rate in Table 12 are taken from Table 5. The reduction in the innovator firm-years reduces the average return of the innovator portfolio for the overall period by 0.27 % points from 0.74 to 0.47 %. This is traceable to the lower return in the first sub-period because the data preprocessing produced a marginal higher innovator return in the second sub-period. The average return of the control firm portfolio for the overall period was a modest 0.21 %. The control firm portfolio has achieved a high return in the first sub-period, but this was offset by a large negative return of 1.13 % in the second sub-period.

Table 13 presents the regression results for the Carhart four-factor model for both the innovator and the control firm portfolio. The alpha of the innovator portfolio, which is now reduced in terms of sample size, is significant and 1.34 % for the entire time period. Although positive, the according point estimates for the sub-periods are not statistically different from zero. In contrast, the portfolio of control firms never exhibits an alpha different from zero at conventional significance levels. Interestingly, the control firm portfolio mimics market movements more closely than the innovator portfolio with a market beta close to one in all time periods. This holds also true for all the following control firm matching specifications.

The parameter estimates of the four-factor model with control firms derived by exact industry matching are shown in Table 14. Compared to the country restriction, the industry restriction yields a slightly different sample of innovator firms. However, the alpha for the entire time span remains significant at 1.37 %. Even though none of the alphas of the control firms are statistically different from zero, it should be noted that the point estimates are always positive. What is more, the z-statistic for the overall period proves to be the highest among all the control firm

Table 7 No. of innovators and control firms across industries (country matched)

Industries	No. of innovators	No. of control firms
Basic materials	13	15
Consumer goods	4	26
Consumer services	1	35
Financials	10	50
Healthcare	3	16
Industrials	68	53
Oil and gas	88	6
Technology	16	28
Telecommunications	0	4
Utilities	34	4
Total	237	237

Table 8 No. of innovators and control firms across countries

Industries	No. of innovators	No. of control firms
Denmark	9	9
Finland	4	4
France	6	6
Germany	140	140
Italy	1	1
Netherlands	2	2
Norway	10	10
Spain	13	13
Sweden	2	2
Switzerland	23	23
United Kingdom	27	27
Total	237	237

Table 9 Annual means of innovator (i), control (c), and market (m) portfolio characteristics

Years	Size			BEME			Momentum			No. of firms		
	I	C	M	I	C	M	I	C	M	I	C	M
2002	859	917	1,359	0.2	0.21	1.03	-5.81	-4.68	-3.69	10	10	4,557
2003	471	526	1,285	1.07	1.41	1.2	-6.29	-6.14	-1.37	14	14	4,338
2004	285	123	1,660	0.93	0.89	0.92	2.06	2.22	2.91	13	13	4,235
2005	642	401	2,103	0.76	0.59	0.81	1.74	2.7	1.69	15	15	4,222
2006	693	597	2,221	0.41	0.32	0.71	5	4.95	1.79	25	25	4,257
2007	1,220	1,223	2,702	0.4	0.42	0.68	1.7	1.24	1.94	31	31	4,530
2008	2,127	1,854	3,178	0.41	0.5	0.78	3.42	4.41	1.8	58	58	4,343
2009	648	573	1,451	1.25	1.41	1.96	-5.14	-5.28	-4.87	71	71	4,556

The month December is used to calculate the means. Size stands for the market value of a firm and is denominated in Mio USD

scenarios. This suggests the presence of a positive industry effect and indicate that the sustainability research team of the ZKB managed to identify outperforming companies within booming industries.

In a last robustness check using one-to-one nearest neighbor PSM, we match both by country and industry. The results depicted in Table 15 back up the outperformance of the innovator portfolio. The reduction of the innovator alpha to 1.05 % and the upward shift of the significance level to 10 % can be explained by the substantial reduction of the sample through the caliper matching. In this scenario as well, the alphas of the control firm portfolio are not significant.

Finally, Table 16 presents the alpha distribution of 1,000 resampled control and innovator portfolios based on the Carhart four-factor model. The parameter values indicate that the alpha distributions are very symmetric. The resampling shows that the 5 % percentile of the control portfolio is always negative while the same

Table 10 Annual medians of innovator (I), control (C), and market (M) portfolio characteristics

Years	Size			BEME			Momentum			No. of firms		
	I	C	M	I	C	M	I	C	M	I	C	M
2002	460	465	75	0.14	0.18	0.68	-6.01	-4.34	-2.01	10	10	4,557
2003	76	59	73	0.76	0.62	0.81	-7.5	-3.2	-0.3	14	14	4,338
2004	78	66	107	0.76	0.79	0.64	1.83	1.99	2.77	13	13	4,235
2005	126	91	140	0.53	0.48	0.57	1.69	2.8	1.82	15	15	4,222
2006	244	234	146	0.26	0.28	0.47	4.96	4.31	1.75	25	25	4,257
2007	345	339	177	0.29	0.38	0.45	2.01	1.89	2	31	31	4,530
2008	545	610	204	0.25	0.38	0.46	4.23	3.84	1.76	58	58	4,343
2009	190	201	76	0.64	0.78	1.04	-4.27	-4.88	-4.29	71	71	4,556

The month December is used to calculate the means. Size stands for the market value of a firm and is denominated in Mio USD

Table 11 Mean absolute portfolio difference as compared to the innovator portfolio over the period 2001–2008

	Mean			Median		
	Size	BEME	MOM	Size	BEME	MOM
Control firms	120.38	0.12	0.5	20.13	0.08	1.13
Remaining market	1,126.75	0.33	1.66	144	0.21	2.25

In every year absolute differences are built between control firms and innovators and the market and innovators for Tables 8 and 9. Then the mean over the years is calculated

Table 12 Average monthly returns over time (geometric returns in percent)

	r_{mt}	r_{ft}	Innovators _t	Control firms _t
1/2002–4/2009	0.30 (0.59)	0.31 (0.01)	0.47 (1.18)	0.21 (0.81)
1/2002–8/2005	1.37 (0.67)	0.23 (0.00)	0.85 (1.34)	1.56 (0.89)
9/2005–4/2009	-0.78 (0.95)	0.39 (0.02)	0.09 (1.96)	-1.13 (1.34)

Average geometric returns are displayed and standard errors are shown in parenthesis. Value weighted returns are presented for the market return r_{mt} and equally weighted returns are displayed for the innovator portfolio. r_{ft} is the risk-free rate

percentile for the innovator portfolio is always substantially above zero for the innovators. Therefore, 90 % of the alphas as enclosed by the 5 and 95 % percentiles are substantially above zero for the innovators in each matching scenario while the same distributions for the control firm portfolio include zero. If the asset pricing model we are using were a perfect description of expected returns for the analyzed portfolios, we would expect a mean of zero for the alpha of the control portfolios.

Table 13 Exact country match: parameter estimates and z-statistics in the four-factor model for different time periods

	Alpha	$r_{mt}-r_{it}$	SMB _t	HML _t	WML _t	R ²
01/2002–04/2009						
Innovators	1.34** (2.17)	1.39*** (12.04)	0.51 (1.52)	−1.12** (−2.43)	−0.33 (−1.53)	0.62
Control firms	−0.18 (−0.47)	0.94*** (8.82)	0.46** (2.07)	0.25 (0.64)	−0.17 (−0.77)	0.57
01/2002–08/2005						
Innovators	1.30 (1.03)	1.30*** (4.94)	1.12** (2.12)	−1.61*** (−2.74)	−0.53* (−1.74)	0.49
Control firms	0.14 (0.15)	0.83*** (5.10)	0.31 (0.55)	0.20 (0.36)	−0.19 (−0.52)	0.27
09/2005–04/2009						
Innovators	1.32 (1.49)	1.38*** (9.13)	0.27 (0.88)	1.01 (1.10)	−0.08 (−0.31)	0.76
Control firms	−0.15 (−0.31)	0.96*** (7.00)	0.51** (2.54)	0.41 (0.81)	−0.18 (−1.06)	0.78

* (**, ***) means that the parameter is different from zero at the 10 % (5 %, 1 %) significance level, respectively. The parentheses below the point estimates contain the z-statistics, corrected for heteroskedasticity and autocorrelation up to three lags

Table 14 Exact industry match: parameter estimates and Z-statistics in the four-factor model for different time periods

	Alpha	$r_{mt}-r_{it}$	SMB _t	HML _t	WML _t	R ²
01/2002–04/2009						
Innovators	1.37** (2.21)	1.40*** (12.25)	0.52 (1.51)	−1.12** (−2.41)	−0.32 (−1.49)	0.62
Control firms	0.78 (1.36)	1.08*** (10.30)	0.61** (2.56)	0.47 (0.74)	−0.46** (−2.07)	0.65
01/2002–08/2005						
Innovators	1.43 (1.13)	1.31*** (5.06)	1.10** (2.09)	−1.63*** (−2.77)	−0.52* (−1.73)	0.50
Control firms	0.92 (0.84)	0.96*** (4.45)	0.78 (1.10)	0.56 (0.72)	−0.80*** (−3.07)	0.50
09/2005–04/2009						
Innovators	1.27 (1.42)	1.37*** (8.86)	0.29 (0.87)	1.10 (1.11)	−0.06 (−0.25)	0.76
Control firms	0.44 (0.66)	1.16*** (7.27)	0.45** (2.39)	1.02 (1.51)	−0.00 (−0.01)	0.78

* (**, ***) means that the parameter is different from zero at the 10 % (5, 1 %) significance level, respectively. The parentheses below the point estimates contain the z-statistics, corrected for heteroskedasticity and autocorrelation up to three lags

Table 15 Exact country-industry match: parameter estimates and Z-statistics in the four-factor model for different time periods

	Alpha	$r_{mt}-r_{it}$	SMB _t	HML _t	WML _t	R ²
01/2002–04/2009						
Innovators	1.09* (1.75)	1.28*** (12.30)	0.55* (1.94)	−1.08*** (−2.82)	−0.39** (−2.06)	0.62
Control firms	0.08 (0.14)	1.00*** (8.12)	0.22 (0.88)	0.14 (0.41)	0.06 (0.40)	0.57
01/2002–08/2005						
Innovators	0.82 (0.65)	1.32*** (4.90)	1.03* (1.94)	−1.41** (−2.66)	−0.52* (−1.80)	0.48
Control firms	0.69 (0.79)	0.84*** (3.62)	0.58 (1.56)	−0.23 (−0.61)	−0.10 (−0.54)	0.29
09/2005–04/2009						
Innovators	1.03 (1.23)	1.22*** (8.81)	0.42 (1.50)	0.78 (0.97)	−0.18 (−0.89)	0.77
Control firms	−0.23 (−0.32)	1.04*** (6.49)	−0.04 (−0.17)	1.01 (1.48)	0.24 (1.00)	0.72

* (**, ***) means that the parameter is different from zero at the 10 % (5, 1 %) significance level, respectively. The parentheses below the point estimates contain the z-statistics, corrected for heteroskedasticity and autocorrelation up to three lags

Table 16 Alpha distribution of random samples of innovator and control firms for the period 01/2002–04/2009

Matching characteristics	Portfoliotype	Alphas			
		Mean	p5	p50	p95
Size, BEME, MOM	Innovators	1.42	0.79	1.42	2.06
	Control firms	0.16	−0.44	0.17	0.79
Size, BEME, MOM, country	Innovators	1.47	0.83	1.49	2.07
	Control firms	0.29	−0.31	0.29	0.92
Size, BEME, MOM, sector	Innovators	1.49	0.88	1.50	2.10
	Control firms	0.31	−0.33	0.31	0.99
Size, BEME, MOM, country, sector	Innovators	0.99	0.42	0.99	1.53
	Control firms	0.23	−0.39	0.23	0.84

p5, p50, and p95 are the 5, 50, and 95 % percentiles, respectively

However, these means turn out to be positive with a monthly outperformance ranging between 0.16 and 0.31 %. This indicates that our alpha estimate for the innovator portfolio should be corrected downwards. We conclude that the returns of our equally weighted innovator portfolio outperformed the market from 1/2002 to 4/2009 by at least around 1 % monthly. To sum up, despite the stringent country and

industry matchings, the performance of the innovator portfolio remains robust in all scenarios, providing additional support for our hypothesis.

6 Conclusion

There are no established characteristics that divide firms into an SRI class, let alone a class that reflects a distinct strategic sustainability concept. Instead, many studies apply indices that are specifically built to arrange such a categorization. It stands to reason that within such indices, elaborate as they are, specific performance-driving effects of SRI might get lost in aggregation. In other words, the validity of conclusions with respect to SRI and financial performance hinges on the construction of the SRI class. Our basket of innovators reflects the specific notion of strategic CSR and circumvents this issue. In addition, the synthetic portfolio offers a glimpse at a portfolio selection stage that is not distorted by active portfolio management skills.

The case for positive abnormal returns of this portfolio is given by a hypothesis that rests upon two channels. On the one hand, value-relevant information of strategic CSR may not have been priced in by the market. On the other hand, market disequilibrium caused by rising demand for SRI assets might have produced positive returns via stock price effects. To test the hypothesis of positive abnormal returns, we conduct a portfolio analysis based on a novel comprehensive dataset with pan-European Carhart risk factors. This accounts for the well-known systematic risks in SRI portfolios and eschews mistaken attributions of observed out- or underperformance. The results of the portfolio analysis strongly support the hypothesis of financial outperformance with a stable alpha of 1.30 % per month.

Even so, one might argue that the financial performance driver might be confounded by bad model problems due to firm characteristics. To mitigate this possibility, we construct synthetic benchmark portfolios with similar control firms, additionally accounting for country and industry clusters. The matched portfolios underpin the significant financial outperformance of the innovators. A final prudent resampling approach yields a more conservative innovators alpha of roughly 1 %. In other words, none of the robustness checks qualify the general results.

It is instructive to discuss the implications of both arguments of the hypothesis. Most CSR research has focused on easily quantifiable CSR characteristics for large firms. By now, potentially value-relevant factors embedded within or proxied by these characteristics can be extracted systematically and with reasonable costs thanks to the surge of CSR databases. In contrast, CSR features that need to be processed more comprehensively on a case to case basis, such as strategies, demand skilled expertise and come at a cost, in particular when the focus is on small firms. One can argue that if there is something value-relevant to be found, it is more likely to happen in areas which are tougher to decipher. So any financial outperformance that lures in these niches has to make up for the research costs that are associated with this task. This suggests that one might keep encountering financial outperformance in market niches that are new and unfamiliar. On that notion, our results suggest that strategic CSR for small and innovative firms seems to evoke a market anomaly in terms of systematic outperformance that has yet to be priced in by the market.

The mere growth of SRI assets as the second argument draws attention to an overlooked dynamic in the existing CSR literature. Whenever investors deviate from market weights, the ensuing rising demand can temporarily give rise to abnormal returns due to stock market frictions. We present empirical evidence to support this argument for Europe in the last decade. With an increasingly saturated SRI market in the US, the ongoing SRI expansion in Europe lends itself to be a more fertile ground for positive abnormal returns, particularly with respect to small stocks. By and large, existing SRI research has focused on big stocks. However, ever since Merton (1987), there is a theoretical case for small stocks, and given our results, it seems worthwhile to engage in future research with European SRI data to raise more evidence. Notably for our portfolio, policy regulations could amplify abnormal returns. For example, the extensive subsidies for renewable energies in Germany could have distorted the allocation of SRI assets.

Our results are of high practical importance for both investors and a firm's management. At first glance, it seems that investing in small and innovative firms that focus on strategic sustainability kills two birds with one stone. SRI in strategic CSR can be quite lucrative, at least on the revenue side. But this comes at a cost. On the expenditure side, research costs have to be taken into account, costs that are likely to be associated with expertise that gradually builds up over time. For the management, our results may serve as important advice. If value-relevant information is disclosed more transparently, the firm is more attractive to the market. It falls to the management how this can be conveyed credibly.

It would be revealing to disentangle the two channels. In our case, a longer time horizon might allow for a distinction of the two effects if we were to observe how differently, if at all, the abnormal behavior evolves in the long run. A convergence of future abnormal returns towards zero would make for a case of errors of investors' expectations. On the other hand, if abnormal returns eventually turn negative, it would suggest that equilibrium in the SRI investor share has been reached. Finally, if both abnormal and absolute returns turn negative, we might observe a shrinking proportion of SRI investors. One might label this situation as the burst of a financial SRI bubble. Given the magnitude of our results, the two channels might well be simultaneously at work. If anything, our data is only weakly indicative in the sense that the lack of an observable negative trend of the alpha point estimates in the rolling regression does not suggest a decreasing effect of SRI growth. Analyzing according US data from the past two decades in light of the slowing growth of SRI assets could shed more light on this question.

Acknowledgments We would like to thank the editor, two anonymous referees, Tommy Lundgren, and Andreas Schrimpf for insightful comments. We are extremely grateful to Peter Steffen Schmidt for providing access to and support with the financial database and helpful comments. We thank the ZKB for financial, infrastructural, and data support, in particular René Nicolodi and his team. Financial support from the Swiss National Science Foundation (SNSF) and from the Commission for Technology and Innovation of the Swiss Confederation (CTI) is gratefully acknowledged.

References

- Aggarwal R, Erel I, Stulz R, Williamson R (2009) Differences in governance practices between US and foreign firms: measurement, causes, and consequences. *Rev Financ Stud* 22(8):3131–3169
- Ambec S, Lanoie P (2007) When and why does it pay to be green? Série Scientifique Scientific Series Montréal Septembre 2007. Working paper
- Bagnoli M, Watts SG (2003) Selling to socially responsible consumers: competition and the private provision of public goods. *J Econ Manag Strategy* 12(3):419–445
- Barber BM, Lyon JD (1997) Detecting long-run abnormal stock returns: the empirical power and specification of test statistics. *J Financ Econ* 43(3):341–372
- Baron DP (2001) Private politics, corporate social responsibility, and integrated strategy. *J Econ Manag Strategy* 10(1):7–45
- Bauer R, Koedijk K, Otten R (2005) International evidence on ethical mutual fund performance and investment style. *J Bank Finan* 29(7):1751–1767
- Bauer R, Derwall J, Otten R (2007) The ethical mutual fund performance debate: new evidence from Canada. *J Bus Ethics* 70(2):111–124
- Bebchuk LA, Cohen A, Ferrell A (2009) What matters in corporate governance? *Rev Finan Stud* 22(2):783–827
- Bebchuk LA, Cohen A, Wang Charles CY (2011) Learning and the disappearing association between governance and returns. *J Financ Econ* (forthcoming)
- Besley T, Ghatak M (2007) Retailing public goods: the economics of corporate social responsibility. *J Public Econ* 91(9):1645–1663
- Capelle-Blancard G, Laguna MA (2010) How does the stock market respond to chemical disasters? *J Environ Econ Manag* 59(2):192–205
- Carhart M (1997) On persistence in mutual fund performance. *J Finan* 52(1):57–82
- Cortez MC, Silva F, Areal N (2012) Socially responsible investment in the global market: the performance of US and European Funds. *Int J Finan Econ* 17(3):254–271
- Dasgupta S, Laplante B, Mamingi N (2001) Pollution and capital markets in developing countries. *J Environ Econ Manag* 42(3):310–335
- DeMiguel V, Garlappi L, Uppal R (2009) Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Rev Finan Stud* 22(5):1915–1953
- Derwall J, Guenster N, Bauer R, Koedijk K (2005) The eco-efficiency premium puzzle. *Finan Anal J* 61(2):51–63
- Derwall J, Koedijk K, Ter Horst J (2011) A tale of values-driven and profit-seeking social investors. *J Bank Finan* 35:2137–2147
- Drucker S, Puri M (2005) On the benefits of concurrent lending and underwriting. *J Finan* 60(6):2763–2799
- Eccles RG, Serafeim G, Krzus MP (2011a) Market interest in nonfinancial information. *J Appl Corp Finan* 23(4):113–127
- Eccles RG, Ioannou I, Serafeim G (2011b) The impact of a corporate culture of sustainability on corporate behavior and performance. Harvard Bus School Working Paper 12–035
- Eurosif (2003) Socially responsible investment among European institutional investors. Retrieved from http://www.eurosif.org/images/stories/pdf/eurosif_srreprt_2003_all.pdf
- Eurosif (2010) European SRI Study. Retrieved from http://www.eurosif.org/images/stories/pdf/Research/Eurosif_2010_SRI_Study.pdf
- Fama EF (1970) Efficient capital markets: a review of theory and empirical work. *J Finan* 25(2):383–417
- Fama EF (1991) Efficient capital markets: II. *J Finan* 46(5):1575–1617
- Fama EF (1998) Market efficiency, long-term returns, and behavioral finance. *J Financ Econ* 49:283–306
- Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J Financ Econ* 33(1):3–56
- FNG (2011) Marktbericht Nachhaltige Geldanlagen 2011: Deutschland, Österreich und die Schweiz. Retrieved from http://www.forum-ng.org/images/stories/nachhaltige_geldanlagen/FNG_Marktbericht_Nov2011_web.pdf
- Friedman M (1970) The social responsibility of business is to increase its profits. *The New York Times Magazine* (13 September)
- Gompers P, Ishii JL, Metrick A (2003) Corporate governance and equity prices. *Quart J Econ* 118(1):107–155

- Greene WH (2002) *Econometric analysis*, 5th edn. Prentice Hall, New Jersey
- Guo SY, Fraser DMW (2009) *Propensity score analysis: statistical methods and applications*, 1st edn. Sage Publications Inc, Thousand Oaks
- Hamilton JT (1995) Pollution as news: media and stock market reactions to the toxics release inventory data. *J Environ Econ Manag* 28(1):98–113
- Heckman J, Navarro-Lozano S (2004) Using matching, instrumental variables, and control functions to estimate economic choice models. *Rev Econ Stat* 86(1):30–57
- Ho DE, Imai K, King G, Stuart EA (2007) Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal* 15(3):199–236
- Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: implications for stock market efficiency. *J Finan* 48(2):65–91
- Karpoff JM, Lott JE Jr, Wehrly EW (2005) The reputational penalties for environmental violations: empirical evidence. *J Law Econ* 48:653–675
- Kempf A, Osthoff P (2007) The effect of socially responsible investing on portfolio performance. *Euro Finan Manag* 13(5):908–922
- Leuven E, Sianesi B (2003) PSMATCH2: stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Retrieved from <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Li X, Zhao X (2006) Propensity score matching and abnormal performance after seasoned equity offerings. *J Empir Finan* 13(3):351–370
- Linn J (2010) The effect of cap-and-trade programs on firms' profits: evidence from the nitrogen oxides budget trading program. *J Environ Econ Manag* 59(1):1–14
- Lundgren T, Olsson R (2010) Environmental incidents and firm value: international evidence using a multi-factor event study framework. *Appl Finan Econ* 20(16):1293–1307
- Lyon JD, Barber BM, Tsai CL (1999) Improved methods for tests of long-run abnormal stock returns. *J Finan* 54(1):165–201
- Markowitz H (1952) Portfolio selection. *J Finan* 7(1):77–91
- McWilliams A, Siegel D (2001) Corporate social responsibility: a theory of the firm perspective. *Acad Manag Rev* 26(1):117–127
- Merton RC (1987) A simple model of capital equilibrium with incomplete information. *J Finan* 62(3):483–510
- Mitchell ML, Stafford E (2000) Managerial decisions and long-term stock price performance. *J Bus* 73(3):287–329
- Mollet JC, Ziegler A (2012) Is socially responsible investing really beneficial? New empirical evidence for the US and European stock markets. MAGKS Discussion Paper No. 28-2012
- Moskowitz MR (1972) Choosing socially responsible stocks. *Bus Soc Rev* 1(1):71–75
- Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708
- Petajisto A (2009) Why do demand curves for stocks slope down? *J Finan Quan Anal* 44(5):1013–1044
- Porter ME, Kramer MR (2006) Strategy and society: the link between competitive advantage and corporate social responsibility. *Harvard Bus Rev* 84(12):78–92
- Porter ME, Kramer MR (2011) Creating shared value. *Harvard Bus Rev* 89(1/2):62–77
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55
- Schmidt PS, Von Arx U, Schrimpf A, Wagner AF, Ziegler A (2011) On the construction of common size, value, and momentum factors in international stock markets: a guide with applications. Working paper
- Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under condition of risk. *J Finan* 19(3):461–488
- Siegel DS, Vitalino DF (2007) An empirical analysis of the strategic use of corporate social responsibility. *J Econ Manag Strategy* 16(3):773–792
- Tirole J (2001) Corporate governance. *Econometrica* 69(1):1–35
- US-Sif (2010) Report on socially responsible investing trends in the United States. Retrieved from <http://ussif.org/resources/research/documents/2010TrendsES.pdf>
- Wall LD (1995) Some Lessons from Basic Finance for Effective Socially Responsible Investing. *Econ Rev—Federal Reserve Bank of Atlanta* 80(1):1–12
- Zurich Cantonal Bank (2011) Fokus Nachhaltigkeit. Retrieved from: https://www.zkb.ch/etc/ml/repository/prospekte_und_broschueren/corporate/studien/218387_fokus_nachhaltigkeit_pdf.File.pdf